Discussion Papers In Economics And Business

Self-selection and Earnings of Migrants:
Evidence from Rural China

Zheren WU

Discussion Paper 08-25

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Self-selection and Earnings of Migrants:
Evidence from Rural China

Zheren WU

Discussion Paper 08-25

June 2008

この研究は「大学院経済学研究科・経済学部記念事業」
基金より援助を受けた、記して感謝する。

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Self-selection and Earnings of Migrants: Evidence from Rural China*

Zheren Wu†

Abstract

Using data from a rural household survey in China, this paper explores the link between employment choice (nonworking, local farm work, local nonfarm work and migratory work) and migrant earnings. We find significant self-selection in migration. Youths, males, better-educated individuals and those in good health are more likely to migrate. In terms of unobserved characteristics, we find positive selection in migration as related to the alternatives of not working and local farm work, and insignificant self-selection as related to local nonfarm work. Controlling for self-selection, the wage returns to gender (male), education and health are lower than those obtained from simple ordinary least squares (OLS), and the returns to experience are higher. More importantly, we find different self-selection between individuals who have moved as pioneers and migrants from households in which other members have already migrated.

JEL classifications: J24; J31; O15

Keywords: Migration, Self-selection, Pioneer migrants

* This paper has benefited from helpful comments provided by Tsunehiro Otsuki, Akira Kousaka, Hisakazu Matsushige, Miki Kohara, Shinya Kajitani and Junyi Shen. I am grateful for financial support from the Matsushita International Foundation and the Setsutaro Kobayashi Memorial Fund.

† Correspondence: Osaka School of International Public Policy, Osaka University; 1-31 Machikaneyama, Toyonaka, Osaka, 5600043, Japan

E-mail: wzheren@osipp.osaka-u.ac.jp

1. Introduction

Migration has long been recognized as a self-selection process. Based on the insights of Roy (1951), Borjas (1987, 1991) investigated how the endogeneity of the migration decision affects labor market performance, in terms of both unobservable characteristics (such as ability, ambition, physical strength and skill) and observable characteristics (such as education, experience, and training). It is expected that individuals compare their potential wages (based on their observable and unobservable human capital) in their home community with those in other regions. Those who migrate have higher expected earnings in other regions than at home and vice versa for those who stay. When migration is an endogenous outcome of optimizing decisions, wage estimates are inconsistent unless corrective measures are taken. This, in turn, will lead to misleading evaluations of policy programs such as public investments in education and subsidies on training (Rosenzweig and Wolpin, 1986; Lanzona, 1998).

The self-selection in migration also generates predictions about how migrants compare with the populations of the origins. The predictions of migrant selectivity are important for policy making. For example, there is a rising concern that large-scale rural-to-urban migration might harm the rural development in China. If able workers move out of the villages, the lack of competent workers will dampen the development of rural economies, causing rural areas to fall further behind urban areas. In this case, local government may have an incentive to carry out policies to restrict migration.[3] Because migrant attributes such as ability and skills are unobservable, policy makers are always limited in obtaining predications for self-selection by using observable characteristics such as education. However, there is no necessary correlation between the self-selection in terms of observable characteristics and in terms of unobservable characteristics. For instance, it is perfectly possible for migrants to be among the most educated in their home communities but perform poorly in the destination (Ghatak et.al., 1996). Hence, investigating self-selection based on unobserved characteristics will yield important implications.

Heckman (1976, 1979) developed a two-step procedure to correct self-selection bias and infer the unobserved characteristics by checking whether individuals' unobserved characteristics from their choice of whether or not to participate in the activity of interest (e.g., migration) are correlated with the outcome of interest (e.g., migrant earnings). Lee (1983) and Dubin and McFadden (1984) developed the approach to apply to multiple choice. In general, when the unobservable characteristics from the migration decision are consistently associated with a higher (lower) wage rate, we conclude migrants are more (less) productive and positively (negatively) selected.

Although the existing literature has provided deep theoretical insight into, and various econometric methodologies toward understanding, the self-selection in migration, empirical studies to ascertain the impact of unobservable characteristics on the migration decision and earnings estimation are still few in number. It might be because information (especially regarding wages) of migrants is always missing from surveys conducted in their origin. As an alternative, Lanzona (1998) analyzed whether the selectivity of

[3] The effect of migration on the economic of source community also depends on remittances and other factors. See, for example, Rozelle, et.al. (1999), de Brauw and Rozelle (forthcoming), Ghatak et.al. (1996).

migration affects the wage structure estimated for those who stay in rural communities. Using Philippine data and Lee's sample selection model, Lanzona found that there is a negative selection bias in nonmigrants' wage estimate, and that as a result migrants were positively selected.

Axelsson and Westerlund (1998) examined the potential influence of migration on total household real income in Sweden. Employing Heckman's selection bias correction approach, they accounted for the potential correlation between unobserved household characteristics exerting influence on both the decision to migrate and on household income but did not find significant self-selection in migration.

In this paper, we use data from a rural household survey conducted by the Ministry of Agriculture of China in the Sichuan and Anhui provinces to explore the link between employment choice (nonworking, local farm work, local nonfarm work and migratory work) and migrant earnings. It is worth pointing out that, under the unique household registration system in China (see Liu, 2005; Au and Henderson, 2006; etc.), entire rural households rarely relocate, and migrants usually maintain their permanent registration in the home and circulate between their place of origin and destination. These features enable the rural household survey to collect data successfully on every household member, including migrants. Therefore, missing observations for migrants in the sample, which is a typical problem in other countries, does not arise.

In contrast to the traditional approach in existing studies (typically a version of Heckman, 1976, 1979; or Lee, 1983), we employ the selectivity correction methodology developed by Dubin and McFadden (1984) and modified by Bourguignon et al. (2007). This sophisticated method, based on the multinomial logit model including multiple correction terms, allows us not only to attribute a selection bias in the estimation of earnings to the allocation of individuals with better or worse unobserved characteristics in migration, but also to link the selection bias to the allocation of individuals to each other alternative. That is, we are allowed not only to identify what the selection bias is, but also from where the bias stems.

The empirical estimates document significant self-selection in migration. In general, youths, males, better-educated individuals and those in good health are more likely to migrate. In terms of unobserved characteristics, we find positive selection in migration as related to the alternatives of not working and local farm work, and insignificant self-selection as related to local nonfarm work. Controlling for self-selection, the wage returns to gender (male), education and health are lower than that obtained from OLS, and the returns to experience are higher.

More interestingly, when we consider the possibility that pioneer migrants may confer a positive externality on (potential) future migrants, the empirical results based on two subgroups, which are divided by whether or not other household members had moved in the past three years, exhibit striking differences. This suggests that the self-selection in migration is tied significantly to other household members' migration experiences.

This paper is structured as follows. Section 2 presents the econometric framework. Section 3 describes the data and the empirical specification. Section 4 reports the empirical results and Section 5 concludes.

2. Econometric Model: Selection Bias Corrections Based on Multiple Choice

Consider a situation in which an individual makes an employment choice, each individual may select among J mutually exclusive alternatives. These alternatives could for example be: (a) not to work; (b) to be a local farm worker; (c) to be a local nonfarm worker; and (d) to migrate from the home village and work in some other location. Let Y_j^* be the utility attainable for an individual if he/she chooses alternative j. We can write the indirect utility function as:

$$Y_j^* = \mathrm{Z}\gamma_j + \eta_{\mathrm{j}}, \qquad \mathrm{j} = 1,...,\mathrm{J}, \tag{1}$$

where the vector Z represents the maximum set of explanatory variables for all employment alternatives, and η_j is a disturbance term.

A rational individual compares the utility attainable given each alternative and selects the alternative s that yields the highest benefit to him/her, that is:

$$Y_s^* > \max_{j \neq s}(Y_j^*), \qquad s \in (1,..., J).$$

Assume the market wage in the s^{th} alternative is given by:

$$\ln w_s = X_s \beta_s + u_s, \tag{2}$$

where X_s is a vector of exogenous variables that determine logarithmic earnings $(\ln w_s)$, and the disturbance u_s satisfies $E(u_s \mid X, Z) = 0$ and $V(u_s \mid X, Z) = \sigma_s^2$.

If there are unobserved characteristics that affect both individuals' choices and their earnings, then the disturbance η_j in (1) and disturbance u_s in (2) will be correlated such that:

$$\begin{aligned} E(\ln w_s \mid Y_s^* > \max_{j \neq s} Y_j^*) &= X_s\beta_s + E(u_s \mid Y_s^* > \max_{j \neq s} Y_j^*) \\ &= X_s\beta_s + E(u_s \mid \max_{j \neq s}(Z_j\gamma_j + \eta_j - Z_s\gamma_s - \eta_s) < 0) \\ &\neq X_s\beta_s. \end{aligned}$$

The term $E(u_s \mid \max_{j \neq s}(Z_j\gamma_j + \eta_j - Z_s\gamma_s - \eta_s) < 0)$ captures the unobservable characteristics affecting decision making and are persistently correlated with the wages. For instance, able individuals could choose to migrate, and potentially reward their ability with a higher wage rate. Because we cannot observe ability, omitting important variables from the OLS regression produces biased estimates of β_s.

The potential inconsistency requires selection bias correction methods, following the seminal insight of Heckman (1976, 1979). When the choices are multiple, two typical approaches based on the multinomial logit specification were developed by Lee (1983) and Dubin and McFadden (1984, hereafter DMF). Lee's model is based on the assumption that the unobservable determinants of the choice of alternative s against any other alternative correlate with the unobservable determinants of the outcome ($\ln w_s$) in the same direction, and thus the model uses only one correction term to explain $E(u_s \mid \max_{j \neq s}(Z_j\gamma_j + \eta_j - Z_s\gamma_s - \eta_s) < 0)$. DMF make no assumption about this. They use multiple correction terms to control for self-selection in the s^{th} alternative as related to each other alternative. The correlations between u_s and $(\eta_j - \eta_s)$ could be of different signs for different j. As shown by Schmertmann (1994) and Bourguignon et al. (2007), Lee avoids the risk of multicolinearity present in DMF but makes a very strong assumption that does not always hold in empirical studies. If Lee's assumptions did not hold, then the results would be inconsistent. Another virtue of DMF's approach is that it identifies not only the direction of the selection bias, but also where the bias stems from, by linking the selection bias to the allocation of individuals to each other alternative. The formulation of the sample selection model in this paper follows DMF and here we briefly outline DMF's method. A more detailed discussion can be found in Dubin and McFadden (1984) and Bourguignon et al. (2007).

Assuming the disturbances η_j in equation (1) are independent and identically Gumbel distributed (the so-called IIA hypothesis), the specification of (1) leads to the multinomial logit model with:

$$P_k = \frac{\exp(Z\gamma_k)}{\sum_{j=1}^{J}\exp(Z\gamma_j)}, \quad \mathrm{j} = 1,...,\mathrm{J}, \quad \mathrm{k} \in (1,...,J), \tag{3}$$

where P_k is the probability that any alternative k is preferred.

Under DMF's assumption that:

$$E(u_s) = \sigma_s \frac{\sqrt{6}}{\pi} \sum r_j(\eta_j - E(\eta_j)), \quad with \sum r_j = 0, \quad j = 1,...,J,$$

conditional on the alternative s being chosen, estimating the following wage function that includes the correction terms yields consistent estimates of β_s:

$$\ln w_s = X_s\beta_s + \sigma_s \frac{\sqrt{6}}{\pi} \sum_{j \neq s} r_j [\frac{P_j \ln(P_j)}{1 - P_j} + \ln(\mathrm{P_s})] + \mathrm{e_s}, \tag{4}$$

where r_j is the correlation coefficient between disturbances u_s and η_j, and e_s is a residual with asymptotic mean zero.

Bourguignon et al. (2007) showed that when we remove the restriction of $\sum r_j = 0$, equation (4) can be written as:

$$\ln w_s = X_s \beta_s + \sigma_s \frac{\sqrt{6}}{\pi} [\sum_{j \neq s} r_j \left(\frac{P_j \ln(P_j)}{1 - P_j} \right) - r_s \ln(\mathrm{P_s})] + \mathrm{e_s}. \tag{5}$$

Their Monte Carlo experiments showed that the equation (5) variant outperforms the traditional DMF correction method when the restriction is violated and provides very similar results when it holds.

Equations (4) and (5) can be estimated in two steps. In the first step, the polychotomous choice model is estimated by the logit maximum likelihood method (equation (3)). Let $\hat{P}_j$ be the predicted probabilities for P_j, $j = 1,..., J$. In the second step, we substitute $\hat{P}_j$, $j = 1,..., J$ into the selectivity correction terms in equation (4) or (5) and we then estimate the function by OLS. Note that while the second step estimates from DMF are consistent, they have inefficient standard errors because of the two-step nature of the procedure. There can be some efficiency gain from using weighted estimations. The form of heteroskedasticity is detailed in Bourguignon et al. (2007) and we use the bootstrap method.

Because $\sigma_s \neq 0$, we conduct a t-test of whether $\sqrt{6}\sigma_s r_j / \pi = 0$ can be used to test $r_j = 0$. The simple OLS covariance matrix is valid only when all the self-selection correction coefficients equal zero. Furthermore, we notice that the value of $\frac{\mathrm{P_j} \ln(P_j)}{1 - P_j}$ increases with the decrease in $\mathrm{P_j}$. Then a positive (negative) coefficient of $\sqrt{6}\sigma_s r_j / \pi$ indicates a positive (negative) selection because the individuals with better (worse) unobserved endowments choose the given alternative s rather than the respective alternative j. For example, a positive self-selection correction coefficient related to local farm work in the migrants' wage equation highlights the higher wages of individuals who work outside the home village compared with random individuals, because people who are more productive (in terms of unobserved characteristics) prefer migration to local farm work.

3. Data and Model Specification

3.1 The data

The empirical analysis is based on the data from the rural household surveys conducted by the Ministry of Agriculture of China (MAC) at the end of December each year. We use data on the Sichuan and Anhui provinces that covers the period from 2003 to 2006. The Sichuan and Anhui provinces are in western and central China, respectively. Both are predominantly rural and are the two largest exporters of rural labor. Thus, the experiences of rural households in these two provinces may shed light on those of rural households elsewhere in less developed regions in China.

The survey includes approximately 2000 households from 33 villages. The original selection of villages and households used stratified random sampling methods. From 2003, the MAC collected data not only at the household level but also on each household member. Household-level data include information on incomes, landholdings and other aspects of household economic activity. Data on individuals include information about employment, education and demographic characteristics. An outstanding feature of our data is that, under the unique household registration system, entire rural households rarely relocate; migrants usually maintain their permanent registration in the home village and circulate between their places of origin and destination. Therefore, another type of sample selection, i.e., the omission from the survey of migrants, which is a problem that typically arises in surveys in most other countries, does not arise in China. One of the shortcomings of the data is that personal earnings are only available in the migration section, thus we can focus on the migrants' wage equation only.

Because most migrants moved repeatedly, our analysis mainly focuses on 2006 data, however utilizes information such as migration experience in the past three years (2003–2005). We include only laborers in the sample. Following the definition by the National Statistical Bureau of China, a laborer is defined as a female aged between 16 and 55 or a male aged between 16 and 60. Anyone who works 90 days or more during the year but fails to meet the age constraints also belongs to the labor force.[4] Full-time students and individuals who had lost the ability to work are excluded. After accounting for missing observations, the valid sample consists of 4,820 individuals from 1,797 households.

We classify these rural laborers into four categories: (1) nonworking individuals; (2) local farm workers; (3) local nonfarm workers; and (4) out-of-village migrants. Anyone who worked less than 90 days in the labor market in a year is defined as a nonworking individual.[5] A migrant is defined as a person who worked outside his or her home village on at least 30 days during the survey year.[6] Under this definition, 2,121 individuals in our sample migrated and 94.15% of them worked outside their villages on at least 90 days; the mean migration period is 254 days. The rest of the population is divided into local farm workers and local nonfarm workers by self-reported main job.

Table 1 presents the allocation of the labor force over the four selection categories and also by gender, age and schooling cohort. As mentioned above, the Sichuan and Anhui provinces are agricultural provinces and large migrant-exporters; 39.63% of the sample population mainly participated in local farm work, and 44% of the population engaged in migratory work. Males appear to lead females in moving out the village (20.84% higher), however had a significantly lower probability of being a local farm worker (23.37% lower). We observe a perceptible effect of age on the employment choice. Younger persons have a higher rate of moving out. Around 70% of those under 31 years old participated in out-of-village work. Older persons are more likely to be involved in local work. In the five age categories, the highest participation rates in local farm and local nonfarm work both appeared in the cohort of over 35 years old, with rates of 54.71% and 10.83%, respectively. It may be because local work can easily be combined with living in the village and old persons are likely to have more experience and more contacts that are relevant for finding local nonfarm jobs. Another notable point is that age seems negatively related to the proportion of unemployment, with 15.55% in the cohort of 16–20 years old and 5.81% in those more than 35 years old.

[4] When we set the minimum number of days as 1 or 30, our results were similar.
[5] When we set the minimum number of days as 1 or 30, our results were similar.
[6] We also set the minimum number of days of migratory work as 1 or 90. The results were similar.

Table 1 Employment selection: by gender, age and education cohort (2006 MAC surveys)

		Employment categories			
	Total (people)	(1)Nonworking	(2)Local farm	(3)Local nonfarm	(4)Migration
Total	4,820	7.61%	39.63%	8.76%	44%
Gender					
Male	2,635	6.11%	29.03%	11.12%	53.74%
Female	2,185	9.43%	52.40%	5.90%	32.90%
Age categories					
16-20	328	15.55%	8.54%	3.05%	72.87%
21-25	544	9.93%	11.21%	6.62%	72.24%
26-30	528	9.28%	15.91%	6.06%	68.75%
31-35	512	8.59%	28.52%	5.56%	57.23%
36-	2,908	5.81%	54.71%	10.83%	28.65%
Education categories					
0-6	2,213	6.06%	62%	7.50%	24.45%
7-9	2,181	7.66%	22.83%	8.76%	60.75%
10-	426	15.49%	9.39%	15.26%	59.86%

The table also shows significant differences in the employment choice among people with different education endowment. Three main points can be noted. First, migration mainly occurred among better-educated people. The proportion is around 60% for those who had at least seven years of schooling. Second, less-educated people are more likely to engage in local farm work. Of those belonging to the 0–6 years of schooling cohort, 62% engaged in local farm work, far exceeding the participation rate of 22.83% and 9.39% for those belonging to the 7–9 years of schooling cohort and more than 9 years of schooling cohort, respectively. Third, compared with less-educated individuals, better-educated people are more likely to be unemployed. We find that 15.49% of individuals who received at least 10 years of schooling did not participate in the labor market. Educated unemployment has been a noteworthy phenomenon in rural China.

3.2 Empirical model and variables

The two-stage approach requires that there must be different sets of independent variables in the two functions. Following the preceding studies (see, for example, the survey of Greenwood, 1975, 1997; Yap, 1977; Lucas, 1997), the empirical specification is defined as:

$$P_j = f(\text{Network, Migration experience, Age, Age squared, Male, Education, Health,}\\ \text{Number of young dependents, Number of other migrants, Household size,}\\ \text{Landholding, Lagged house value, Village dummies}),\quad j = 1,2,3,4, \qquad (6)$$

$$\begin{aligned}\ln(\text{daily earnings}) = g(&\text{ Netwrok, Migration experience, Age, Age squared, Male,}\\ &\text{Education, Health, Number of young dependents, correction terms), iff } j = 4, \qquad (7)\end{aligned}$$

where (6) is the employment selection function and (7) is the migrants' earnings function.

Table 2 presents definitions of main variables and reports their means and standard deviations for the whole sample and the four categories. We use logarithmic average daily migrant earnings as a dependent variable in the earnings function. The mean value of migrants' earnings per day is 28.8 yuan.

From the discussions above and the migration literature, we expect that those who have migration experience or migration networks, males, better-educated individuals and healthy people are more likely to move and meet with success at their destinations. Note that the networks dummy takes the value 1 if any other household member had been a migrant, while the individual had not moved in the past three years, and takes the value 0 otherwise. This dummy indicates the presence of an effective migration network that provides information to potential migrants without previous migration experience. Older people are less likely to migrate, because they have less time to pay back investments and they often have their family to take care of. However, we predict a positive coefficient on age in the earnings function, as age is an instrument of experience. The squared term is added to account for nonlinearities in the impact of age. For the number of young dependents, individuals with children are generally less eager to move, however the need for a large house or the (prospective) schooling of children might prompt a move (Sandell, 1977; Mincer, 1978; Nivalainen, 2004; etc.). In addition, people with dependent children are usually viewed as being more responsible. Therefore, the effect of the number of young dependents on the migration decision might be ambiguous, while the effect on earnings might be positive.

Variables of the number of other migrants, household size, landholding, lagged house value and village factors are hypothesized to influence the occupation decisions, however have no effect on the migrants' wage in their destinations. The effect of the number of other migrants on an individual's migration decision is unpredictable: when migration occurs by one or more household member, other household members may choose to move with them or to remain in the village to look after the home. Workers belonging to large households are expected to be less mobile because of their greater attachment to the family. Large areas of irrigated land reduce the need for working off-farm and the possibility of unemployment. The variable of lagged house value measures wealth. People from poor households are expected to have an incentive to migrate because they are more motivated than the rich by potential income gains from migration. The highest average lagged house value is found in those who worked in local nonfarm sectors. A set of village dummies is included in (6) to control village characteristics such as village infrastructure, geographic location, economic shocks because of natural disasters, migration networks with neighbors, and local wage level.

Consistent with the statistics in Table 1, Table 2 shows that those who did not participate in the labor market are relatively young. Note also that the average schooling years of those who are unemployed is significantly higher than that of local farm workers, and even slightly higher than that of local nonfarm

Table 2 Definitions and descriptive statistics (means) of selected variables

	Definition	Total	(1) Nonworking	(2) Local farm	(3) Local nonfarm	(4) Migration
Daily earnings	average daily earnings, (yuan)					28.796
						(16.706)
#Network	dummy,=1 if other members had moved in 2003-2005	0.333	0.485	0.568	0.308	0.100
	and the individual had not move in 2003-2005	(0.471)	(0.501)	(0.496)	(0.462)	(0.300)
#Migration experience	dummy,=1 if the individual had moved in	0.482	0.270	0.135	0.258	0.875
	2003-2005	(0.500)	(0.444)	(0.342)	(0.438)	(0.331)
Age	years old	40.559	36.275	48.045	44.600	33.756
		(13.898)	(13.516)	(12.044)	(13.638)	(11.694)
#Male	dummy,=1 if male	0.547	0.439	0.401	0.694	0.668
		(0.498)	(0.497)	(0.490)	(0.461)	(0.471)
Education	years of schooling	6.611	7.275	5.042	7.107	7.811
		(2.990)	(3.446)	(2.806)	(3.117)	(2.340)
Health	=5, very good; =4 good; =3 normal;	4.362	4.365	4.163	4.284	4.556
	=2, bad; =1, very bad.	(0.783)	(0.808)	(0.843)	(0.880)	(0.643)
Number of young dependents	Number of preschool children and students (people)	0.814	0.774	0.833	0.867	0.793
		(0.807)	(0.747)	(0.859)	(0.862)	(0.756)
Number of other migrants	(people)	1.045	0.973	1.033	0.611	1.156
		(0.988)	(0.989)	(0.997)	(0.878)	(0.977)
Household size	(people)	4.427	4.738	4.348	4.393	4.452
		(1.530)	(1.537)	(1.660)	(1.679)	(1.359)
Landholding	Per capita landholding (mu)	5.224	5.579	5.741	4.397	4.861
		(4.601)	(4.580)	(5.109)	(4.175)	(4.124)
Lagged house value	value of house in 2005 (1,000 yuan)	26.783	33.138	25.034	37.341	25.157
		(34.575)	(70.181)	(26.089)	(37.568)	(30.648)
N		4,820	367	1,910	422	2,121

Notes: Standard errors in brackets; # indicate dummy variable.

workers. Why did young educated people not participate in the labor market? A possible explanation for this phenomenon is that these individuals' reservation wages are relatively high, and they did not want to work as a farmer. If the opportunity of local nonfarm employment is limited or the wages are low, they cannot or will not enter the local nonfarm sector. However, why did they not work outside the village? A potential reason might be that they fear an uncertain life in other locations because of a lack of information. From our data, we find that for individuals in the households in which other members had not moved in the past three years, 33.85% of 16–20-year-old young laborers and 20.98% of those who have received at least 10 years of schooling were unemployed. The proportions are significantly lower for individuals from households in which other members had migration experience, with 11.03% and 12.72%, respectively.

According to Carrington et al. (1996) and McKenzie and Rapoport (2007), the presence of pioneer migrants may lower moving costs and enhance migration job prospects of other household members (potential future migrants). They increase the information available to other members and reduce the risks of moving by sending information about how to migrate, where to look for work, what wages to expect, and information on migration costs and risks and how to reduce them. Furthermore, if future migrants move to the same region, pioneer migrants can support these subsequent migrants at their destinations by, for example, providing them with job search assistance, helping them to find housing, and even by extending credit, providing lodgings and in other ways lowering the psychological costs of leaving home. Hence, individuals from households in which other members have had a migration experience are more likely to migrate and less likely to be a nonworker. They are even less likely to be a local wage earner, if migration returns are high. To address this issue, we divide the full sample into two subgroups according to whether or not other members had moved in the past three years, and see whether choice making and the possible self-selection of migrants differs between those who moved as pioneers and those having access to a network of previous migrants in the household.[7]

4. Empirical Results

4.1. Multinomial logit results on employment selection

In the first step of our DMF analysis, we estimate the employment choice model as given in equation (6). This step not only provides insights into the determinants of the choice of an individual, but also generates bias correction terms for the second step. Table 3 reports the average marginal effects from the multinomial logit model for the full sample and the two subgroups. In each case, the test based on a chi-square statistic proposed by Hausman and McFadden (1984) does not reject the null hypothesis of IIA at the 99% significance level, indicating that distinguishing between nonworking, local farm work, local nonfarm work and migration satisfies the basic assumption in DMF.

[7] Migrants from households where other members had not moved in the past three years can be viewed as pioneer migrants. However, because of data limitations we cannot rule out the possibility that the migrants who come from households in which other members have migration experiences were in fact the first to move and therefore are pioneers. Nevertheless, we investigate the effects of prior migrants.

Table 3 Marginal effects from multinomial logit for employment decision

	Full sample				*Subgroup A*: Others had NOT moved in 2003-2005				*Subgroup B*: Others had moved in 2003-2005			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Nonworking	Local farm	Local nonfarm	Migration	Nonworking	Local farm	Local nonfarm	Migration	Nonworking	Local farm	Local nonfarm	Migration
#Network	-0.011	-0.027 *	-0.035 **	0.073 **								
	(0.010)	(0.015)	(0.008)	(0.016)								
#Migration experience	-0.116 **	-0.366 **	-0.101 **	0.583 **	-0.091 **	-0.329 **	-0.135 **	0.555 **	-0.108 **	-0.308 **	-0.04 **	0.456 **
	(0.008)	(0.019)	(0.008)	(0.026)	(0.009)	(0.026)	(0.015)	(0.033)	(0.007)	(0.014)	(0.007)	(0.017)
Age	-0.004 **	0.021 **	0.001	-0.018 **	-0.01 **	0.012 **	0.011 **	-0.012 **	-0.002	0.025 **	-0.003 **	-0.019 **
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
Age squared/100	0.001	-0.015 **	0.0005	0.014 **	0.008 **	-0.007	-0.011 **	0.009 **	0.0004	-0.019 **	0.005 **	0.014 **
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)
#Male	-0.018 **	-0.129 **	0.042 **	0.104 **	-0.024 **	-0.129 **	0.069 **	0.084 **	-0.01	-0.128 **	0.03 **	0.108 **
	(0.007)	(0.011)	(0.010)	(0.011)	(0.012)	(0.022)	(0.022)	(0.020)	(0.009)	(0.013)	(0.010)	(0.013)
Education	0.002	-0.013 **	0.008 **	0.004 *	0.005 **	-0.024 **	0.016 **	0.003	0.001	-0.008 **	0.004 **	0.004
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Health	-0.031 **	0.011	-0.006	0.026 **	-0.016	0.014	-0.017	0.019	-0.034 **	0.008	-0.004	0.03 **
	(0.005)	(0.007)	(0.006)	(0.008)	(0.010)	(0.016)	(0.014)	(0.014)	(0.007)	(0.008)	(0.006)	(0.009)
Number of young	-0.024 **	0.021 **	-0.007	0.01	-0.021 **	0.025 *	-0.016	0.012	-0.02 **	0.02 **	-0.003	0.003
	(0.006)	(0.008)	(0.006)	(0.008)	(0.009)	(0.014)	(0.012)	(0.012)	(0.008)	(0.010)	(0.007)	(0.011)
Number of other migrants	-0.015 **	0.036 **	-0.029 **	0.008	-0.002	0.014	-0.06 **	0.049 **	-0.017 **	0.03 **	-0.018 **	0.006
	(0.005)	(0.007)	(0.006)	(0.007)	(0.019)	(0.028)	(0.026)	(0.019)	(0.005)	(0.008)	(0.005)	(0.008)
Household size	0.021 **	-0.024 **	0.016 **	-0.012 **	0.023 **	-0.034 **	0.023 **	-0.012	0.018 **	-0.021 **	0.008 *	-0.006
	(0.004)	(0.006)	(0.004)	(0.005)	(0.007)	(0.010)	(0.009)	(0.009)	(0.005)	(0.007)	(0.005)	(0.007)
Landholding	-0.001	0.008 **	-0.006 **	-0.001	-0.003	0.014 **	-0.012 **	0.001	-0.001	0.007 **	-0.004 *	-0.002
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Lagged house value	0.0002 **	-0.000 *	0.0003 **	-0.000	0.0001	-0.000	0.001 **	-0.001 *	0.0002 **	-0.000	0.0002 **	-0.000
	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000	(0.000
Log likelihood	-3051.				-841.3				-2111.			
LR test [degrees of	4860.9 **	[132]			1685.4 **	[129]			3229.1 **	[129]		
McFadden's Adj R2	0.443				0.5				0.433			
N	4,820				1,399				3,421			

Notes: Standard errors in brackets; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively; # indicate dummy variable. Village dummies are all controlled.

In the results for the full sample in the first four columns, we find that both the individual's and other household members' migration experience (network) increase the probability of current migration. At the same time, these migration experiences influence other alternatives negatively, although the negative effect of networks on the alternative of being a nonworking person does not have a high significance level. As has been inferred from the descriptive statistics, age has a negative effect on the probability of leaving the home village. At the same time, younger people show a greater propensity for unemployment. As for gender and education, males and better-educated individuals are more likely to participate in local nonfarm work and migration. Health has a positive effect on the likelihood of migration and a negative effect on nonworking. The average marginal effects of other control variables are as expected, and are thus consistent with the theory and with existing empirical findings.

Columns 5 to 12 in Table 3 present the results on the two subgroups, with columns 5 to 8 indicating individuals from households in which other members had not moved in the past three years (subgroup A), and columns 9 to 12 indicating those belonging to households in which one or more other members had moved (subgroup B). The table reveals prominent differences between the two subgroups. The most noteworthy of the results are as follows.

First, for the choice of being a nonworker, as seen in subgroup A (column 5), without a network of prior migrants in the household, age, gender and education all significantly affect the likelihood of not working at least at the 5% significance level. Generally, youths, females and better-educated individuals are more likely to be unemployed. However, in subgroup B, with at least one other household member having migration experience (column 6), these variables all failed to affect the likelihood of unemployment at the 10% statistical significance level. There is no strong evidence to support the hypotheses that individuals with different age, gender and education endowment have different probabilities of unemployment.

Second, as for the likelihood of being a local nonfarm worker, for individuals belonging to households in which other members did not have migration experience (subgroup A, column 8) there is an inverted U-shaped effect of age; however for individuals from households in which other members had migration experience (subgroup B, column 11), the effect of age is horseshoe shaped. A possible explanation is that, because of the lack of networks for individuals in subgroup A, local nonfarm work is a relatively favorable selection, because older laborers have more experience and more contacts that are relevant for finding local nonfarm jobs, age increases the possibility of being a local nonfarm worker at first. When people reach late middle-age, the effect of age becomes negative, indicating a link between nonfarm work and retirement. However because of the presence of migration networks in subgroup B, facing a lower moving cost and risk, youths, especially those of marriageable age, are more eager to move out to make potential income gains from migration, and saving money for marriage or the arrival of children. When children go to school, location ties are stronger and they may prefer to return to the village. In line with this explanation, we observed a positive effect of age on the probability of local nonfarm work after individuals reach their 30s.

Third, in subgroup A, in the absence of a household member with migration experience, better-educated workers seem to choose rural nonfarm work over migration. The average marginal effect of education on the likelihood of choosing local nonfarm work is 0.016 and significant at the 1% level, while it is 0.003 on the likelihood of migration and insignificant at the 10% level. The null hypothesis that education affects the likelihood of the two alternatives equally is rejected at the 10% significance level. In subgroup B in which at least one other household members had migrated in the past three years, the average marginal effects of education on the two selections are very similar, reported by the same values of 0.004, at the 5% significance level for local nonfarm work and at the 12% significance level for migration. The hypothesis that education has the same effect on the two alternatives is accepted at greater than the 95% level in the case of subgroup B.

Fourth, with the absence of a pioneer migrant in the household, in subgroup A, we fail to find significant effects attributable to health. Health is not an important determinant in individuals' employment decisions. On the other hand, in subgroup B, for people belonging to households in which other members had experienced migration, the average marginal effect of health on the likelihood of migration is positive (0.03), and the effect on the likelihood of nonworking is negative (-0.034); both are reported at the 1% significance level. People in good health are more likely to migrate and are less likely to be unemployed.

These differences indicate that the presence of people who have experienced migration significantly changes other members' employment choices. Compared with those belonging to households where other members had experienced migration (subgroup B), in households without a prior migrant (subgroup A), health concerns are not a main reason of unemployment; youths and females are more likely to face unemployment; moreover, in the absence of such a network, better-educated people are more likely to engage in local nonfarm work rather than migrate. If these better-educated laborers fail to find nonfarm jobs in the village, they prefer to wait and not work. From the estimation, the mean value of the predicted probability of unemployment in subgroup A is 21.66% for individuals between 16 to 20 years old, 10.15% for females, and 17.39% for those who received 10 or more years' schooling; the corresponding probability in subgroup B is significantly lower[8] at 10.74%, 9.22% and 9.09%, respectively. Hence, we conclude that prior migrants play an active role in promoting migration and reducing youth unemployment and female unemployment in rural regions; they also enhance efficiency of the utilization of human resources by reducing educated unemployment.

4.2. Estimation of migrants' earnings

In the second step, we estimate the migrants' earnings equation. To explore the importance of self-selection and highlight the differences between the OLS and the selection-corrected wage estimates, we report both the OLS and DMF results. Because the selection bias corrections based on equation (5) are always preferred to the approach based on equation (4) (Bourguignon et al., 2007), and in our case the results of the two specifications were slightly different, we report the results based on equation (5) only.[9]

[8] Significant at the 1%–11% level in the one-sided t tests.

[9] The Wu-Hausman test cannot reject the null hypothesis that there is no difference between the two specifications (without bootstrap) at the 10% level. The results based on equation (4) are available upon request.

Table 4 presents the estimated results for logarithmic average daily earnings of migrants, where Table 4-2 includes cross terms of education and the dummy of migration experience into the specification in Table 4-1. The estimator variances for DMF are all bootstrapped with 100 replications to deal with the heteroskedasticity.[10] In each DMF specification, we find significant coefficients on the selection bias correction terms, which indicate that there is significant self-selection in migration and the simple OLS estimators are inconsistent and DMF is preferred.[11]

Using the full sample, results contained in column 4 show that the coefficients on the selection bias correction terms related to the nonworking sector and the local farm sector are both positive and significant at the 5% level. In other words, migrant earnings are upward biased because the individuals with better unobserved characteristics are more likely to choose migration rather than to be unemployed or to work in the local farm sector. The coefficient on the correction term as related to local nonfarm work is negative but insignificant at the 10% level, indicating that there is no significant difference in productivity based on unobserved characteristics between migrants and local nonfarm workers.

Confirming the self-selection in the two subgroups (columns 5 and 6), in each case we find a positive coefficient on the correction term as related to local farm work (at least at the 10% significance level). Similar to the results for the full sample, there is an upward bias in migrants' earnings because of people who are well-endowed in term of unobserved characteristics prefer migration to local farm work. However, as observed in the coefficients on the other two correction terms, whether other household members do or do not have migration experience makes a great difference. First, consider the correction term related to the nonworking sector. The coefficient is small (0.09) and statistically insignificant at the 10% level in subgroup A (other members without migration experience), while the corresponding coefficient is 0.386 and significant at the 1% level in subgroup B (other members had moved in the past three years). Another difference is seen in the coefficient on the correction term related to the local nonfarm sector. The coefficient is negative (–0.377) and statistically significant at the 5% level for migrants belonging to subgroup A;[12] however, the coefficient is positive and insignificant at the 10% level for those belonging to subgroup B. In the absence of any other household member who migrated (subgroup A), there is no significant self-selection in migration as related to unemployment, but negative selection as related to local nonfarm work. In other words, individuals with the best unobserved characteristics prefer local nonfarm work; those with the worst endowment choose to work in the local farm sector; the remaining workers who lie in the middle-level of unobserved endowments choose migration or unemployment. However, in the presence of a member with migration experience (subgroup B), migrants are positively selected relative to unemployed individuals; and statistically insignificantly selected relative to local nonfarm workers. That is, the local nonfarm sector and migration sector attract workers with higher productivity, while less able

[10] We also repeated the bootstrap algorithm 50, 200 and 1000 times. The results were similar.
[11] The Wu-Hausman test rejects all the null hypotheses of no difference between the estimates of OLS and DMF (without bootstrap) at the 10% significance level.

[12] Lee's assumptions do not hold in the case of subgroup A because the direction of self-selection in migration as related to local farm work and that as related to local nonfarm work are different.

Table 4 Estimated earning equation results

	4-1						4-2					
	OLS			DMF			OLS			DMF		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full	Subgroup A	Subgroup B	Full	Subgroup A	Subgroup B	Full	Subgroup A	Subgroup B	Full	Subgroup A	Subgroup B
#Network	0.060			0.064			0.057			0.044		
	(0.079)			(0.092)			(0.079)			(0.093)		
#Migration experience	0.090	0.127	0.028	-0.095	-0.009	-0.200 ***	0.003	-0.150	-0.012	-0.279	-0.630 **	-0.287 *
	(0.072)	(0.078)	(0.039)	(0.110)	(0.150)	(0.075)	(0.100)	(0.280)	(0.110)	(0.190)	(0.280)	(0.150)
Age	0.041 ***	0.044 ***	0.038 ***	0.048 ***	0.041 ***	0.045 ***	0.041 ***	0.044 ***	0.038 ***	0.047 ***	0.042 ***	0.045 ***
	(0.006)	(0.013)	(0.007)	(0.006)	(0.011)	(0.008)	(0.006)	(0.013)	(0.007)	(0.007)	(0.014)	(0.007)
Age squared/100	-0.047 ***	-0.050 ***	-0.043 ***	-0.053 ***	-0.046 ***	-0.048 ***	-0.047 ***	-0.051 ***	-0.043 ***	-0.052 ***	-0.047 ***	-0.047 ***
	(0.008)	(0.016)	(0.010)	(0.007)	(0.012)	(0.010)	(0.008)	(0.016)	(0.010)	(0.008)	(0.015)	(0.009)
#Male	0.244 ***	0.247 ***	0.237 ***	0.183 ***	0.176 ***	0.187 ***	0.244 ***	0.250 ***	0.238 ***	0.184 ***	0.177 ***	0.187 ***
	(0.025)	(0.071)	(0.027)	(0.024)	(0.059)	(0.030)	(0.025)	(0.071)	(0.027)	(0.030)	(0.061)	(0.034)
Education	0.018 ***	0.038 ***	0.013 **	0.012 **	0.023	0.010 *						
	(0.005)	(0.012)	(0.006)	(0.006)	(0.014)	(0.006)						
Education*New							0.009	0.007	0.009	-0.004	-0.044	0.001
							(0.011)	(0.033)	(0.012)	(0.012)	(0.033)	(0.013)
Education*Repeat							0.020 ***	0.042 ***	0.014 **	0.016 **	0.030 **	0.012 *
							(0.006)	(0.012)	(0.007)	(0.007)	(0.015)	(0.007)
Health	0.106 ***	0.103 ***	0.107 ***	0.095 ***	0.113 ***	0.087 ***	0.106 ***	0.101 **	0.107 ***	0.096 ***	0.110 **	0.088 ***
	(0.018)	(0.040)	(0.021)	(0.021)	(0.037)	(0.022)	(0.018)	(0.040)	(0.021)	(0.019)	(0.045)	(0.021)
Number of young dependents	0.066 ***	0.081 **	0.059 ***	0.063 ***	0.089 **	0.062 ***	0.066 ***	0.079 **	0.060 ***	0.063 ***	0.085 **	0.062 ***
	(0.016)	(0.033)	(0.018)	(0.014)	(0.037)	(0.018)	(0.016)	(0.033)	(0.018)	(0.017)	(0.033)	(0.019)
DMF1 (related to nonworking)				0.337 ***	0.090	0.386 ***				0.305 ***	0.028	0.366 ***
				(0.120)	(0.180)	(0.130)				(0.097)	(0.200)	(0.110)
DMF2 (related to local farm work)				0.306 **	0.354 *	0.278 *				0.276 **	0.353 *	0.259 *
				(0.130)	(0.190)	(0.140)				(0.120)	(0.210)	(0.160)
DMF3 (related to local nonfarm work)				-0.108	-0.377 **	0.160				-0.127	-0.477 **	0.153
				(0.120)	(0.170)	(0.130)				(0.130)	(0.220)	(0.120)
DMF4 (related to migration)				0.028	0.054	0.041				0.055	0.092	0.056
				(0.037)	(0.066)	(0.047)				(0.048)	(0.069)	(0.053)
Constant	1.511 ***	1.296 ***	1.668 ***	1.782 ***	1.638 ***	1.977 ***	1.642 ***	1.552 ***	1.703 ***	1.932 ***	2.199 ***	2.049 ***
	(0.160)	(0.320)	(0.160)	(0.200)	(0.350)	(0.180)	(0.160)	(0.410)	(0.180)	(0.220)	(0.430)	(0.170)
Adjusted R-squared	0.12	0.13	0.1	0.12	0.15	0.11	0.12	0.13	0.1	0.12	0.15	0.11
N	2,121	467	1,654	2,121	467	1,654	2,121	467	1,654	2,121	467	1,654

Notes: Standard errors in brackets; ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively; # indicate dummy variable.

Subgroup A refers individuals who come from households in which other members had not moved between 2003 and 2005;

Subgroup B refers individuals who come from households in which other members had moved between 2003 and 2005;

DMF_j, $j=1,2,3,4$ is the coefficient on correction terms of $P_j \ln P_j/(1-P_j)$, $j=1,2,3$ and $\ln P_4$, respectively; 100 bootstrap replications are performed in DMF.

workers either do not work or are located in the local farm sector. The different self-selection in the two subgroups indicates that employment choice is significantly tied to other members' migration experience. The presence of such a migration network promotes an efficient allocation of labor resources in that people with better unobserved characteristics flow out of the nonworking sector, or perhaps even the local nonfarm sector, to the migration sector.

The differences in the estimated coefficients between OLS and DMF are equally interesting. In each case, OLS overestimates the gender difference and the return to education. This is the result of the positive selection in migration in terms of gender (male), education and unobserved characteristics. The difference on the return to experience (age) and health between OLS and DMF provides another set of interesting observations. OLS underestimates the return to experience and overestimates the return to health for the full sample and subgroup B. It is understandable: migrants are negatively selected in terms of experience (age) and positively selected in health, because the self-selection based on unobserved characteristics is positive on average. These factors lead to a downward bias in the estimator of the return to experience (age) and an upward bias in the return to health when we fail to correct for self-selection bias. However, for subgroup A, OLS slightly overestimates the return to experience and underestimates the return to health. Recall that both experience and health affect the likelihoods of migration and local nonfarm work in the opposite direction. The observation reflects that the positive selection in migration as related to local farm work is offset by the negative selection as related to local nonfarm work, and indicates that the latter selectivity is somewhat dominant.

In general, the selection bias corrected results show that migrants' wages are concave in age. Daily earnings increase with age at a decreasing rate; they reach a maximum in the mid-40s, and start decreasing thereafter. We find that there are significant gender differences in wages, with men earnings about 18% more than women. Education seems to makes a positive difference to earnings, although the coefficient on education in the estimation of subgroup A was insignificant. With regard to other variables, good health is rewarded in the labor market by higher earnings. Young dependents provide an incentive to increase earnings. We are surprised to find the coefficient on the dummy of migration experience is negative and significant at the 1% level in subgroup B. Perhaps the most plausible explanation for this observation is that repeat migrants spend more time in leisure after they grow into their new environment, and this negative effect on work hours was dominant, offsetting the expected positive effect of the higher wage rate.[13]

In order to see the effect of migration experience on the wage rate more clearly, we introduce cross terms of migration experience and education in our specification. Movers who had migration experience in the past three years are defined as repeat migrants, and those who had not moved in the past three years are defined as new migrants. The results are contained in Table 4-2.[14] We find that education does not affect new migrants' earnings significantly, however has a positive effect on repeat migrants' earnings. It

[13] Because of data limitations, we cannot control for work hours.

[14] Correction terms here are obtained from a multinomial logit model with cross terms of migration experience and education. The results of the multinomial logit model are similar to those reported in Table 3-3, and thus we do not present them here. These results are available upon request.

indicates that education enhances migrants' earnings by increasing their skill through on-the-job experience over time. It also provides some evidence that migration experience has a positive effect on the wage rate.

More interestingly, we find different returns to education in the two subgroups. Each additional year of schooling significantly (at the 5% significance level) increases the earnings by 3% for pioneer migrants; and the marginal effect of education for migrants from households in which some other members had moved is 1.2% (at the 10% significance level). The result of a one-sided t test verifies that the return to education is higher for those who moved as pioneers than for migrants from households in which other members had moved (at the 1% significance level). The finding is consistent with the hypothesis as we had mentioned that other members' migration experience might confer a positive externality on later migrants (for example, by helping later migrants find a good paid job) and thus education is more important to those without networks and those who moved as pioneers. Results for other variables and selective terms are very similar to those in Table 4-1.

5. Conclusion

In this paper, we addressed the question of self-selection in migration and how it affects migrants' earnings. Using data from a rural household survey conducted by the Ministry of Agriculture of China in Sichuan and Anhui provinces, and utilizing the two-step selection bias correction method developed by Dubin and McFadden (1984) and modified by Bourguignon et al. (2007), we explored the link between employment choice (nonworking, local farm work, local nonfarm work and migratory work) and migration earnings. We found that migrants in rural China do not make up a random sample of the population from the villages of origin. Our multinomial logit results suggest that youths, males, the better-educated, and healthy individuals are more likely to participate in migratory work. Moreover, earnings results are substantively distorted by selection bias as related to unobservable human capital. In particular, simple OLS overestimated the gender difference and the return to schooling, which may lead to false expectations on migration, for example, providing a disincentive to migrate among females and less-educated persons; or even misleading policy making and policy evaluation such as public schooling investments.

Unlike most literature on this topic, this study not only attributes a selection bias in the estimation of earnings to the allocation of individuals with better or worse unobserved characteristics in migration, but also links the selection bias to the allocation of individuals to each other alternative. In the full sample, we find positive selection in migration against local farm work and nonworking sector and insignificant selection against local nonfarm work. It indicates that the more capable individuals (in terms of unobservable characteristics) choose to participate in migration and local nonfarm work, and those who are less able are left in the local farm sector or remain out of work.

More interestingly, when we see the results on the two subgroups (whether other members had or had not migration experience in the past three years), we find striking differences between them. For migrants from households where other members had not moved (thus making the individual a pioneer), there is no significant self-selection as related to nonworking, but a negative selection as related to local nonfarm work. However, for migrants from households where other members had moved, positive self-selection is reported as related to the alternative of not to work; no significant self-selection is found as related to local nonfarm work. These differences tell us that the presence of individuals with migration experience in the household facilitates migration and promotes an efficient allocation of labor resources, especially in that people with better unobserved characteristics flow out of unemployment to the migration sector.

Our findings lead to several important implications. First, migration enhances efficiency. It reallocates individuals, especially more productive individuals, from low productivity (such as unemployment and local farm work) to high productivity activities. Second, potential income gains appear to provide an incentive to migrate among individuals with better endowments. However, in contrast to the simple pattern of positive self-selection in migration found in the literature, our results suggest that there is insignificant or even negative self-selection in migration in terms of both education and unobserved characteristics as related to the local nonfarm sector. The local nonfarm sector still attracts the most productive laborers. It eases the concern that rural-to-urban migration might impede industrial development in rural China. Third, although the migration network does not significantly contribute to the reduction of total rural unemployment, building public networks to provide rural workers with information about jobs and living conditions at destinations would have a multiplier effect on the reduction in educated unemployment. Furthermore, it promotes employment of youths and females (who are disadvantaged in the labor market). Finally, we would expect government subsidies for education and pre-migration training such as the "Sunshine Project"[15] to enhance migrants' earnings at their destinations, although the returns might be overvalued in existing studies. In addition, such programs particularly benefit those without networks and those who moved as pioneers.

Because the empirical results and implications are based on data from less developed provinces in China, we should be cautious about making generalizations about other countries and relatively richer regions in China. In addition, the question of where migrants move to is of great relevance, however, is not addressed here. Future work should look more closely into the relationship between moving distance and self-selection.

References

Au, C.-C., Henderson, J.V., 2006. How migration restrictions limit agglomeration and productivity in China. Journal of Development Economics 80, 350–388.

Axelsson, R., Westerlund O., 1998. A panel study of migration, self-selection and household real income. Journal of Population Economics 11,112–126.

Borjas, G.J., 1987. Self-selection and the earnings of immigrants, American Economic Review 77, 531–553.

[15] In March 2004, the central government launched the 'Sunshine Project' to train new rural migrants in their home areas, with a concentration on the poorest of these areas. Millions of rural laborers receive benefit from this program each year.

Borjas, G.J., 1991. Immigration and self-selection, in Abowd and Freeman, 27–96

Bourguignon, F., Fournier, M., Gurgand, M., 2007. Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. Journal of Economic Surveys, 21, 174–205.

Carrington W. J., Detragiache E., Vishwanath T., 1996. Migration with endogenous moving costs. American Economic Review 86, 909–930.

de Brauw, A., Rozelle, S., Migration and household investment in rural China. China Economic Review, forthcoming.

Dubin, J.A., McFadden, D., 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52, 345-362.

Ghatak S., Levine, P., Price, S.W., 1996. Migration theories and evidence: an assessment. Journal of Economic Surveys 10, 159–198.

Greenwood, M.J., 1975. Research on internal migration in the United States: A survey. Journal of Economic Literature 13, 397–433.

Greenwood, M.J., 1997. Internal migration in developed countries, In Handbook of Population and Family Economics, edited by M. Rosenzweig and O. Stark. Vol. 1B. Amsterdam: North Holland.

Hausman, J., McFadden, D., 1984. Specifieation test for the multinomial logit model. Econometrica 52, 1219–1240.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5, 475–492.

Heckman, J.J., 1979. Sample bias as a specification error. Econometrica 47, 153–161.

Lanzona, L. A., 1998. Migration, self-selection and earnings in Philippine rural communities. Journal of Development Economics 56, 27–50.

Lee, L.F., 1983. Generalized econometric models with selectivity. Econometrica 51, 507–512.

Liu, Z., 2005. Institutions and inequality: The Hukou System in China. Journal of Comparative Economics 33, 133–157

Lucas, R.E.B., 1997. Internal migration in developing countries. In Handbook of Population and Family Economics, edited by M. Rosenzweig and O. Stark. Vol. 1B. Amsterdam: North Holland.

Mckenzie, D., Rapoport, H., 2007. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. Journal of Development Economics 84, 1–24

Mincer, J., 1978. Family migration decisions. Journal of Political Economy 86, 749–773.

Nivalainen, S., 2004. Determinants of family migration: short moves vs. long moves. Journal of Population Economics 17, 157–175.

Rosenzweig, M.R., Wolpin K.I., 1986. Evaluating the effects of optimally distributed public programs: child health and family planning interventions. American Economic Review 76, 470–482.

Roy, A.D., 1951. Some thought on the distribution of earnings. Oxford Economic Papers 3, 135–146.

Rozelle, S., Taylor, E., de Brauw, A., 1999. Migration, remittance, and agricultural productivity in China. American Economic Review 89, 287–291.

Sandell, S.H. 1977. Women and the economics of family migration. The Review of Economics and Statistics 59, 768–782.

Schmertmann, C.P. 1994. Selectivity bias correction methods in polychotomous sample selection models. Journal of Econometrics 60, 101–132.

Yap, L.Y.L., 1977. The attraction of cities: a review of the migration literature. Journal of Development Economics 4, 239–264.